PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61N 43/04, C12Q 1/68, C12N 15/00, C07H 21/02, A61K 39/00

A1

(11) International Publication Number: WO 97/47358

(43) International Publication Date: 18 December 1997 (18.12.97)

(21) International Application Number: PCT/US97/09884

(22) International Filing Date: 6 June 1997 (06.06.97)

(30) Priority Data:

60/020,494	11 June 1996 (11.06.96)	US
9614731.9	12 July 1996 (12.07.96)	GB
60/033,534	20 December 1996 (20.12.96)	US

(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(71) Applicant *(for all designated States except US)*: MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DONNELLY, John, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LIU, Margaret, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(74) Common Representative: MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).

(54) Title: SYNTHETIC HEPATITIS C GENES

(57) Abstract

This invention relates to novel methods and formulations of nucleic acid pharmaceutical products, specifically formulations of nucleic acid vaccine products and nucleic acid gene therapy products.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE OF THE INVENTION
SYNTHETIC HEPATITIS C GENES

CROSS-REFERENCE TO RELATED APPLICATIONS
Not applicable.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D
Not applicable.

REFERENCE TO MICROFICHE APPENDIX
Not applicable.

FIELD OF THE INVENTION
Not applicable.

BACKGROUND OF THE INVENTION

This invention relates to novel nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

Hepatitis C Virus

Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH).

Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma. Thus, a need exists for an effective method for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in M. A. Brinton, in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologics are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the in situ generation of the protein in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

Therefore, this invention contemplates methods for introducing nucleic acids into living tissue to induce expression of proteins. The invention provides a method for introducing viral proteins into the antigen processing pathway to generate virus-specific immune responses including, but not limited to, CTLs. Thus, the need for specific therapeutic agents capable of eliciting desired prophylactic immune responses against viral pathogens is met for HCV virus by this invention. Of particular importance in this therapeutic approach is the ability to induce T-cell immune responses which can prevent infections even of virus strains which are heterologous to the strain from which the antigen gene was obtained. Therefore, this invention provides DNA constructs encoding viral proteins of the hepatitis C virus core, envelope (E1), nonstructural (NS5) genes or any other HCV genes which encode products which generate specific immune responses including but not limited to CTLs.

DNA Vaccines

Benvenisty, N., and Reshef, L. [PNAS **83**, 9551-9555, (1986)] showed that $CaCl_2$-precipitated DNA introduced into mice intraperitoneally (i.p.), intravenously (i.v.) or intramuscularly (i.m.) could be expressed. The i.m. injection of DNA expression vectors without $CaCl_2$ treatment in mice resulted in the uptake of DNA by the muscle cells and expression of the protein encoded by the DNA . The plasmids were maintained episomally and did not replicate. Subsequently, persistent expression has been observed after i.m. injection in skeletal muscle of rats, fish and primates, and cardiac muscle of rats. The technique of using nucleic acids as therapeutic agents was reported in WO90/11092 (4 October 1990), in which polynucleotides were used to vaccinate vertebrates.

It is not necessary for the success of the method that immunization be intramuscular. The introduction of gold microprojectiles coated with DNA encoding bovine growth hormone (BGH) into the skin of mice resulted in production of anti-BGH antibodies in the mice. A jet injector has been used to transfect skin, muscle, fat, and mammary tissues of living animals. Various methods for introducing nucleic acids have been reviewed. Intravenous injection of a DNA:cationic liposome complex in mice was shown by Zhu et al., [Science 261:209-211 (9 July 1993) to result in systemic expression of a cloned transgene. Ulmer et al., [Science 259:1745-1749, (1993)] reported on the heterologous protection against influenza virus infection by intramuscular injection of DNA encoding influenza virus proteins.

The need for specific therapeutic and prophylactic agents capable of eliciting desired immune responses against pathogens and tumor antigens is met by the instant invention. Of particular importance in this therapeutic approach is the ability to induce T-cell immune responses which can prevent infections or disease caused even by virus strains which are heterologous to the strain from which the antigen gene was obtained. This is of particular concern when dealing with HIV as this virus has been recognized to mutate rapidly and many virulent isolates have been identified [see, for example, LaRosa et al.,

Science 249:932-935 (1990), identifying 245 separate HIV isolates]. In response to this recognized diversity, researchers have attempted to generate CTLs based on peptide immunization. Thus, Takahashi et al., [Science 255:333-336 (1992)] reported on the induction of broadly cross-reactive cytotoxic T cells recognizing an HIV envelope (gp160) determinant. However, those workers recognized the difficulty in achieving a truly cross-reactive CTL response and suggested that there is a dichotomy between the priming or restimulation of T cells, which is very stringent, and the elicitation of effector function, including cytotoxicity, from already stimulated CTLs.

Wang et al. reported on elicitation of immune responses in mice against HIV by intramuscular inoculation with a cloned, genomic (unspliced) HIV gene. However, the level of immune responses achieved in these studies was very low. In addition, the Wang et al., DNA construct utilized an essentially genomic piece of HIV encoding contiguous Tat/*REV*-gp160-Tat/*REV* coding sequences. As is described in detail below, this is a suboptimal system for obtaining high-level expression of the gp160. It also is potentially dangerous because expression of Tat contributes to the progression of Karposi's Sarcoma.

WO 93/17706 describes a method for vaccinating an animal against a virus, wherein carrier particles were coated with a gene construct and the coated particles are accelerated into cells of an animal.

The instant invention contemplates any of the known methods for introducing polynucleotides into living tissue to induce expression of proteins. However, this invention provides a novel immunogen for introducing proteins into the antigen processing pathway to efficiently generate specific CTLs and antibodies.

Codon Usage and Codon Context

The codon pairings of organisms are highly nonrandom, and differ from organism to organism. This information is used to construct and express altered or synthetic genes having desired levels of translational efficiency, to determine which regions in a genome are protein coding regions, to introduce translational pause sites into

heterologous genes, and to ascertain relationship or ancestral origin of nucleotide sequences

The expression of foreign heterologous genes in transformed organisms is now commonplace. A large number of mammalian genes, including, for example, murine and human genes, have been successfully inserted into single celled organisms. Standard techniques in this regard include introduction of the foreign gene to be expressed into a vector such as a plasmid or a phage and utilizing that vector to insert the gene into an organism. The native promoters for such genes are commonly replaced with strong promoters compatible with the host into which the gene is inserted. Protein sequencing machinery permits elucidation of the amino acid sequences of even minute quantities of native protein. From these amino acid sequences, DNA sequences coding for those proteins can be inferred. DNA synthesis is also a rapidly developing art, and synthetic genes corresponding to those inferred DNA sequences can be readily constructed.

Despite the burgeoning knowledge of expression systems and recombinant DNA, significant obstacles remain when one attempts to express a foreign or synthetic gene in an organism. Many native, active proteins, for example, are glycosylated in a manner different from that which occurs when they are expressed in a foreign host. For this reason, eukaryotic hosts such as yeast may be preferred to bacterial hosts for expressing many mammalian genes. The glycosylation problem is the subject of continuing research.

Another problem is more poorly understood. Often translation of a synthetic gene, even when coupled with a strong promoter, proceeds much less efficiently than would be expected. The same is frequently true of exogenous genes foreign to the expression organism. Even when the gene is transcribed in a sufficiently efficient manner that recoverable quantities of the translation product are produced, the protein is often inactive or otherwise different in properties from the native protein.

It is recognized that the latter problem is commonly due to differences in protein folding in various organisms. The solution to this problem has been elusive, and the mechanisms controlling protein folding are poorly understood.

The problems related to translational efficiency are believed to be related to codon context effects. The protein coding regions of genes in all organisms are subject to a wide variety of functional constraints, some of which depend on the requirement for encoding a properly functioning protein, as well as appropriate translational start and stop signals. However, several features of protein coding regions have been discerned which are not readily understood in terms of these constraints. Two important classes of such features are those involving codon usage and codon context.

It is known that codon utilization is highly biased and varies considerably between different organisms. Codon usage patterns have been shown to be related to the relative abundance of tRNA isoacceptors. Genes encoding proteins of high versus low abundance show differences in their codon preferences. The possibility that biases in codon usage alter peptide elongation rates has been widely discussed. While differences in codon use are associated with differences in translation rates, direct effects of codon choice on translation have been difficult to demonstrate. Other proposed constraints on codon usage patterns include maximizing the fidelity of translation and optimizing the kinetic efficiency of protein synthesis.

Apart from the non-random use of codons, considerable evidence has accumulated that codon/anticodon recognition is influenced by sequences outside the codon itself, a phenomenon termed "codon context." There exists a strong influence of nearby nucleotides on the efficiency of suppression of nonsense codons as well as missense codons. Clearly, the abundance of suppressor activity in natural bacterial populations, as well as the use of "termination" codons to encode selenocysteine and phosphoserine require that termination be context-dependent. Similar context effects have been shown to influence the fidelity of translation, as well as the efficiency of translation initiation.

Statistical analyses of protein coding regions of E. coli have demonstrate another manifestation of "codon context." The presence of a particular codon at one position strongly influences the frequency of occurrence of certain nucleotides in neighboring codons, and these context constraints differ markedly for genes expressed at high versus low levels. Although the context effect has been recognized, the predictive value of the statistical rules relating to preferred nucleotides adjacent to codons is relatively low. This has limited the utility of such nucleotide preference data for selecting codons to effect desired levels of translational efficiency.

The advent of automated nucleotide sequencing equipment has made available large quantities of sequence data for a wide variety of organisms. Understanding those data presents substantial difficulties. For example, it is important to identify the coding regions of the genome in order to relate the genetic sequence data to protein sequences. In addition, the ancestry of the genome of certain organisms is of substantial interest. It is known that genomes of some organisms are of mixed ancestry. Some sequences that are viral in origin are now stably incorporated into the genome of eukaryotic organisms. The viral sequences themselves may have originated in another substantially unrelated species. An understanding of the ancestry of a gene can be important in drawing proper analogies between related genes and their translation products in other organisms.

There is a need for a better understanding of codon context effects on translation, and for a method for determining the appropriate codons for any desired translational effect. There is also a need for a method for identifying coding regions of the genome from nucleotide sequence data. There is also a need for a method for controlling protein folding and for insuring that a foreign gene will fold appropriately when expressed in a host. Genes altered or constructed in accordance with desired translational efficiencies would be of significant worth.

Another aspect of the practice of recombinant DNA techniques for the expression by microorganisms of proteins of industrial and pharmaceutical interest is the phenomenon of "codon

preference". While it was earlier noted that the existing machinery for gene expression is genetically transformed host cells will "operate" to construct a given desired product, levels of expression attained in a microorganism can be subject to wide variation, depending in part on specific alternative forms of the amino acid-specifying genetic code present in an inserted exogenous gene. A "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. That these forms provide the message for only 20 different amino acids (as well as transcription initiation and termination) means that some amino acids can be coded for by more than one codon. Indeed, some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon. For reasons not completely understood, alternative codons are not at all uniformly present in the endogenous DNA of differing types of cells and there appears to exist a variable natural hierarchy or "preference" for certain codons in certain types of cells.

As one example, the amino acid leucine is specified by any of six DNA codons including CTA, CTC, CTG, CTT, TTA, and TTG (which correspond, respectively, to the mRNA codons, CUA, CUC, CUG, CUU, UUA and UUG). Exhaustive analysis of genome codon frequencies for microorganisms has revealed endogenous DNA of E. coli most commonly contains the CTG leucine-specifying codon, while the DNA of yeasts and slime molds most commonly includes a TTA leucine-specifying codon. In view of this hierarchy, it is generally held that the likelihood of obtaining high levels of expression of a leucine-rich polypeptide by an E. coli host will depend to some extent on the frequency of codon use. For example, a gene rich in TTA codons will in all probability be poorly expressed in E. coli, whereas a CTG rich gene will probably highly express the polypeptide. Similarly, when yeast cells are the projected transformation host cells for expression of a leucine-rich polypeptide, a preferred codon for use in an inserted DNA would be TTA.

The implications of codon preference phenomena on recombinant DNA techniques are manifest, and the phenomenon may

serve to explain many prior failures to achieve high expression levels of exogenous genes in successfully transformed host organisms-a less "preferred" codon may be repeatedly present in the inserted gene and the host cell machinery for expression may not operate as efficiently. This phenomenon suggests that synthetic genes which have been designed to include a projected host cell's preferred codons provide a preferred form of foreign genetic material for practice of recombinant DNA techniques.

Protein Trafficking

The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence.

Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct destinations. Over the past few years the dissection of the molecular machinery for targeting and localization of proteins has been studied vigorously. Defined sequence motifs have been identified on proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins.

SUMMARY OF THE INVENTION

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid products, when introduced directly into muscle cells,

induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the nucleotide sequence of the V1Ra vector.

Figure 2 is a diagram of the V1Ra vector.

Figure 3 is a diagram of the Vtpa vector.

Figure 4 is the VUb vector

Figure 5 shows an optimized sequence of the HCV core antigen.

Figure 6 shows V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb and VUb.HCV1CorePAb.

Figure 7 shows the Hepatitis C Virus Core Antigen Sequence.

Figure 8 shows codon utilization in human protein-coding sequences (from Lathe et al.).

Figure 9 shows an optimized sequence of the HCV E1 protein.

Figure 10 shows an optimized sequence of the HCV E2 protein.

Figure 11 shows an optimized sequence of the HCV E1 +E2 proteins.

Figure 12 shows an optimized sequence of the HCV NS5a protein.

Figure 13 shows an optimized sequence of the HCV NS5b protein.

DETAILED DESCRIPTION OF THE INVENTION

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH). Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma. Thus, a need exists for an effective method for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in M. A. Brinton, in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologics are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the generation of the encoded protein in situ in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

The standard techniques of molecular biology for preparing and purifying DNA constructs enable the preparation of the DNA therapeutics of this invention. While standard techniques of molecular biology are therefore sufficient for the production of the products of this invention, the specific constructs disclosed herein provide novel therapeutics which surprisingly produce cross-strain

protection, a result heretofore unattainable with standard inactivated whole virus or subunit protein vaccines.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1 µg to 1 mg, and preferably about 10 µg to 300 µg is administered directly into muscle tissue. Subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, or inhalation delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided.

The DNA may be naked, that is, unassociated with any proteins, adjuvants or other agents which impact on the recipients immune system. In this case, it is desirable for the DNA to be in a physiologically acceptable solution, such as, but not limited to, sterile saline or sterile buffered saline. Alternatively, the DNA may be associated with surfactants, liposomes, such as lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture, (see for example WO93/24640) or the DNA may be associated with an adjuvant known in the art to boost immune responses, such as a protein or other carrier. Agents which assist in the cellular uptake of DNA, such as, but not limited to, calcium ions, detergents, viral proteins and other transfection facilitating agents may also be used to advantage. These agents are generally referred to as transfection facilitating agents and as pharmaceutically acceptable carriers. As used herein, the term gene refers to a segment of nucleic acid which encodes a discrete polypeptide. The term pharmaceutical, and vaccine are used interchangeably to indicate compositions useful for inducing immune responses. The terms construct, and plasmid are used interchangeably. The term vector is used to indicate a DNA into which genes may be cloned for use according to the method of this invention.

The following examples are provided to further define the invention, without limiting the invention to the specifics of the examples.

EXAMPLE 1

V1J EXPRESSION VECTORS:

V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment.

pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. We removed the entire *lac* operon from this vector, which was unnecessary for our purposes and may be detrimental to plasmid yields and heterologous gene expression, by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase , treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in E. coli and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The ampicillin resistance marker was replaced with the neomycin resistance marker to yield vector V1Jneo.

An Sfi I site was added to V1Jneo to facilitate integration studies. A commercially available 13 base pair Sfi I linker (New England BioLabs) was added at the Kpn I site within the BGH sequence of the vector. V1Jneo was linearized with Kpn I, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt Sfi I linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with Sfi I) was comparable to expression of the same genes in V1Jneo (with Kpn I).

Vector V1Ra (Sequence is shown in Figure 1; map is shown in Figure 2) was derived from vector V1R, a derivative of the V1Jns vector. Multiple cloning sites (*Bgl*II, *Kpn*I, *Eco*RV, *Eco*RI, *Sal*I, and *Not*I) were introduced into V1R to create the V1Ra vector to improve the convenience of subcloning. V1Ra vector derivatives containing the tpa leader sequence and ubiquitin sequence were generated (Vtpa (Figure 3) and Vub (Figure 4), respectively). Expression of viral antigen from Vtpa vector will target the antigen protein into the exocytic pathway, thus producing a secretable form of the antigen proteins. These secreted proteins are likely to be captured by professional antigen presenting cells, such as macrophages and dendritic cells, and processed and presented by class II molecules to activate CD4+ Th cells. They also are more likely to efficiently simulate antibody responses. Expression of viral antigen through VUb vector will produce a ubiquitin and antigen fusion protein. The uncleavable ubiquitin segment (glycine to alanine change at the cleavage site, Butt et al., JBC 263:16364, 1988) will target the viral antigen to ubiquitin-associated proteasomes for rapid degradation. The resulting peptide fragments will be transported into the ER for antigen presentation by class I molecules. This modification is attempted to enhance the class I molecule-restricted CTL responses against the viral antigen (Townsend et al, JEM 168:1211, 1988).

EXAMPLE 2
DESIGN AND CONSTRUCTION OF THE SYNTHETIC GENES

A. Design of Synthetic Gene Segments for HCV Gene Expression:

Gene segments were converted to sequences having identical translated sequences (except where noted) but with alternative codon usage as defined by R. Lathe in a research article from *J. Molec. Biol.* Vol. 183, pp. 1-12 (1985) entitled "Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data: Theoretical and Practical Considerations". The methodology described below was based on our hypothesis that the known inability to express a gene efficiently in mammalian cells is a consequence of the overall transcript composition. Thus, using alternative codons encoding the same protein sequence may remove the constraints on HCV gene expression. Inspection of the codon usage within HCV genome revealed that a high percentage of codons were among those infrequently used by highly expressed human genes. The specific codon replacement method employed may be described as follows employing data from Lathe et al.:

1. Identify placement of codons for proper open reading frame.

2. Compare wild type codon for observed frequency of use by human genes (refer to Table 3 in Lathe et al.).

3. If codon is not the most commonly employed, replace it with an optimal codon for high expression based on data in Table 5.

4. Inspect the third nucleotide of the new codon and the first nucleotide of the adjacent codon immediately 3'- of the first. If a 5'-CG-3' pairing has been created by the new codon selection, replace it with the choice indicated in Table 5.

5. Repeat this procedure until the entire gene segment has been replaced.

6. Inspect new gene sequence for undesired sequences generated by these codon replacements (e.g., "ATTTA" sequences,

inadvertent creation of intron splice recognition sites, unwanted restriction enzyme sites, etc.) and substitute codons that eliminate these sequences.

7. Assemble synthetic gene segments and test for improved expression.

B. HCV CORE ANTIGEN SEQUENCE

The consensus core sequence of HCV was adopted from a generalized core sequence reported by Bukh et al. (PNAS, 91:8239, 1994). This core sequence contains all the identified CTL epitopes in both human and mouse. The gene is composed of 573 nucleotides and encodes 191 amino acids. The predicted molecular weight is about 23 kDa.

The codon replacement was conducted to eliminate codons which may hinder the expression of the HCV core protein in transfected mammalian cells in order to maximize the translational efficiency of DNA vaccine. Twenty three point two percent (23.2%) of nucleotide sequence (133 out of 573 nucleotides) were altered, resulting in changes of 61.3% of the codons (117 out 191 codons) in the core antigen sequence. The optimized nucleotide sequence of HCV core is shown in Figure 5.

C. CONSTRUCTION OF THE SYNTHETIC CORE GENE

The optimized HCV core gene (Figure 5) was constructed as a synthetic gene annealed from multiple synthetic oligonucleotides. To facilitate the identification and evaluation of the synthetic gene expression in cell culture and its immunogenicity in mice, a CTL epitope derived from influenza virus nucleoprotein residues 366-374 and an antibody epitope sequence derived from SV40 T antigen residues 684-698 were tagged to the carboxyl terminal of the core sequence (Figure 6). For clinical use it may be desired to express the core sequence without the nucleoprotein 366-374 and SV40 T 684-698 sequences. For this reason, the sequence of the two epitopes is flanked by two *Eco*RI sites which will be used to excise this fragment of

sequence at a later time. Thus an embodiment of the invention for clinical use could consist of the V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb, or VUb.HCV1CorePAb plasmids that had been cut with EcoRI, annealed, and ligated to yield plasmids V1Ra.HCV1Core, Vtpa.HCV1Core, and VUb.HCV1Core.

The synthetic gene was built as three separate segments in three vectors, nucleotides 1 to 80 in V1Ra, nucleotides 80 to 347 (*Bst*XI site) in pUC18, and nucleotides 347 to 573 plus the two epitope sequence in pUC18. All the segments were verified by DNA sequencing, and joined together in V1Ra vector.

D. HCV Gene Expression Constructs:

In each case, the junction sequences from the 5' promoter region (CMVintA) into the cloned gene is shown. The position at which the junction occurs is demarcated by a "/", which does not represent any discontinuity in the sequence.

The nomenclature for these constructs follows the convention: "Vector name-HCV strain-gene".

V1Ra.HCV1.CorePAb

---IntA--AGA TCT ACC / ATG AGC--HCV.Core.--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH---

Vtpa.HCV1.CorePAb

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--HCV.Core.--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH---

VUb.HCV1.CorePAb.

---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--HCV.Core.--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH---

V1Ra.HCV1.Core

---IntA--AGA TCT ACC / ATG AGC--HCV.Core.--GCC / TAA A / GTC GAC--BGH---

Vtpa.HCV1.Core

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--HCV.Core.--GCC / TAA A / GTC GAC--BGH---

VUb.HCV1.Core

---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--HCV.Core.--GCC / TAA A / GTC GAC--BGH---

E. OTHER SYNTHETIC HCV GENES

Using similar codon optimization techniques, synthetic genes encoding the HCV E1 (Figure 9), HCV E2 (Figure 10), HCV E1+E2 (Figure 11), HCV NS5a (Figure 12) and HCV NS5b (Figure 13) proteins were created.

WHAT IS CLAIMED:

1. A synthetic polynucleotide comprising a DNA sequence encoding an HCV protein selected from the group consisting of HCV core protein, HCV E1 protein, HCV E1+E2 protein, HCV NS5a protein, HCV NS5b protein and fragments thereof, the DNA sequence comprising codons optimized for expression in a vertebrate host.

2. A plasmid vector comprising the polynucleotide of Claim 1, the plasmid vector being suitable for immunization of a vertebrate host.

3. The polynucleotide of Claim 1 which is HCV genotype I/Ia core.

4. The polynucleotide of Claim 1 having the sequence

```
  1 ATGAGCACCA AcCCcAAgCC cCAgAGgAAg ACCAAgaGgA ACACCAACaG gaGgCCccAG GAtGTgAAGT TcCCtGGgGG  80
 81 aGGcCAGATt GTgGGaGGgG TcTACcTGcT GCCcaGgAGG GGCCCcAGGc TGGGggTGaG gGCtACcagG AAGACcTCtG 160
161 AGaGGTCcCA gCCcaGgGGc AGGaGgCAGC CcATCCCCAA GGCcaGgaGG CCtGAgGGcc GcACcTGGGC cCAgCCtGGc 240
241 TACCCcTGGC CCCTgTATGG CAATGAaGGC TGtGGcTGGG CtGGcTGGCT gCTGTCCCCC aGgGGCTCca GGCCctccTG 320
321 GGGCCCcACa GACCCCaGGa GgaGGTCcaG gAAccTGGGc AAGGTgATtG AcACcCTgAC cTGtGGcTTc GCcGACCTgA 400
401 TGGGcTACAT cCCcCTgGTg GGagCtCCtC TgGGaGGgGC gGCtAGGGCt CTGGCcCATG GgGTgAGGGT gCTGGAgGAt 480
481 GGgGTgAACT ATGCcACtGG cAAccTGCCt GGcTGCTCcT TcTCcATcTT cCTgCTGGCc CTGCTcTCCT GCCTGACagT 560
561 gCCtGCTTcT GCc                                                                          573
```

5. The plasmid vector of Claim 2 having the sequence

```
   1 GATATTGGCT ATTGGCCATT GCATACGTTG TATCCATATC ATAATATGTA CATTTATATT GGCTCATGTC CAACATTACC   80
  81 GCCATGTTGA CATTGATTAT TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT  160
 161 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC CATTGACGTC AATAATGACG  240
 241 TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC  320
 321 AGTACATCAA GTGTATCATA TGCCAAGTAC GCCCCCTATT GACGTCAATG ACGGTAAATG GCCCGCCTGG CATTATGCCC  400
 401 AGTACATGAC CTTATGGGAC TTTCCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA TTACCATGGT GATGCGGTTT  480
 481 TGGCAGTACA TCAATGGGCG TGGATAGCGG TTTGACTCAC GGGGATTTCC AAGTCTCCAC CCCATTGACG TCAATGGGAG  560
 561 TTTGTTTTGG CACCAAAATC AACGGGACTT TCCAAAATGT CGTAACAACT CCGCCCCATT GACGCAAATG GGCGGTAGGC  640
 641 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA GACGCCATCC ACGCTGTTTT  720
 721 GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGCG GCCGGGAACG GTGCATTGGA ACGCGGATTC CCCGTGCCAA  800
 801 GAGTGACGTA AGTACCGCCT ATAGAGTCTA TAGGCCCACC CCCTTGGCTT CTTATGCATG CTATACTGTT TTTGGCTTGG  880
 881 GGTCTATACA CCCCCGCTTC CTCATGTTAT AGGTGATGGT ATAGCTTAGC CTATAGGTGT GGGTTATTGA CCATTATTGA  960
 961 CCACTCCCCT ATTGGTGACG ATACTTTCCA TTACTAATCC ATAACATGGC TCTTTGCCAC AACTCTCTTT ATTGGCTATA 1040
1041 TGCCAATACA CTGTCCTTCA GAGACTGACA CGGACTCTGT ATTTTTACAG GATGGGGTCT CATTTATTAT TTACAAATTC 1120
1121 ACATATACAA CACCACCGTC CCCAGTGCCC GCAGTTTTTA TTAAACATAA CGTGGGATCT CCACGCGAAT CTCGGGTACG 1200
1201 TGTTCCGGAC ATGGGCTCTT CTCCGGTAGC GGCGGAGCTT CTACATCCGA GCCCTGCTCC CATGCCTCCA GCGACTCATG 1280
1281 GTCGCTCGGC AGCTCCTTGC TCCTAACAGT GGAGGCCAGA CTTAGGCACA GCACGATGCC CACCACCACC AGTGTGCCGC 1360
1361 ACAAGGCCGT GGCGGTAGGG TATGTGTCTG AAAATGAGCT CGGGGAGCGG GCTTGCACCG CTGACGCATT TGGAAGACTT 1440
1441 AAGGCAGCGG CAGAAGAAGA TGCAGGCAGC TGAGTTGTTG TGTTCTGATA AGAGTCAGAG GTAACTCCCG TTGCGGTGCT 1520
1521 GTTAACGGTG GAGGGCAGTG TAGTCTGAGC AGTACTCGTT GCTGCCGCGC GCGCCACCAG ACATAATAGC TGACAGACTA 1600
1601 ACAGACTGTT CCTTTCCATG GGTCTTTTCT GCAGTCACCG TCCTTAgatc tgccATGAGC ACCAACCCCA AGCCCCAGAG 1680
1681 GAAGACCAAG AGGAACACCA ACAGGAGGCC CCAGGATGTG AAGTTCCCTG GGGGAGGCCA GATTGTGGGA GGGGTCTACC 1760
1761 TGCTGCCCAG GAGGGGCCCC AGGCTGGGGG TGAGGGCTAC CAGGAAGACC TCTGAGAGGT CCCAGCCCAG GGGCAGGAGG 1840
1841 CAGCCCATCC CCAAGGCCAG GAGGCCTGAG GGCCGCACCT GGGCCCAGCC TGGCTACCCC TGGCCCCTGT ATGGCAATGA 1920
1921 AGGCTGTGGC TGGGCTGGCT GGCTGCTGTC CCCCAGGGGC TCCAGGCCCT CCTGGGGCCC CACAGACCCC AGGAGGAGGT 2000
2001 CCAGGAACCT GGGCAAGGTG ATTGACACCC TGACCTGTGG CTTCGCCGAC CTGATGGGCT ACATCCCCCT GGTGGGAGCT 2080
2081 CCTCTGGGAG GGGCGGCTAG GGCTCTGGCC CATGGGGTGA GGGTGCTGGA GGATGGGGTG AACTATGCCA CTGGCAACCT 2160
2161 GCCTGGCTGC TCCTTCTCCA TCTTCCTGCT GGCCCTGCTC TCCTGCCTGA CAGTGCCTGC TTCTGCCgaa ttcgctccca 2240
2241 atgagaacat ggagaccatg aaccagccct accacatctg ccgcggcttc acctgcttca agaagtaaac ccggaattcc 2320
2321 tcaagtcgaC AGCGGCCGCG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTTGTTTGC CCCTCCCCCG TGCCTTCCTT 2400
2401 GACCCTGGAA GGTGCCACTC CCACTGTCCT TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 2480
```

```
2481 CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA GGATTGGGAA GACAATAGCA GGCATGCTGG GGATGCGGTG 2560
2561 GGCTCTATGG GTACGGCCGC AGCGGCCTTA ATTAAGGCCG CAGGGGGGGT ACCTAGGTGC TGAAGAATTG ACCCGGTTCC 2640
2641 TCGACCGGTA AAAAGGCCGC GTTGCTGGCG TTTTTCCATA GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC 2720
2721 AAGTCAGAGG TGGCGAAACC CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG CGCTCTCCTG 2800
2801 TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA AGCGTGGCGC TTTCTCAATG CTCACGCTGT 2880
2881 AGGTATCTCA GTTCGGTGTA GGTCGTTCGC TCCAAGCTGG GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC 2960
2961 CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT GGTAACAGGA 3040
3041 TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG CCTAACTACG GCTACACTAG AAGGACAGTA 3120
3121 TTTGGTATCT GCGCTCTGCT GAAGCCAGTT ACCTTCGGAA AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC 3200
3201 TGGTAGCGGT GGTTTTTTTG TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT TTGATCTTTT 3280
3281 CTACGTGATC CCGTAATGCT CTGCCAGTGT TACAACCAAT TAACCAATTC TGATTAGAAA AACTCATCGA GCATCAAATG 3360
3361 AAACTGCAAT TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA GAAAACTCAC 3440
3441 CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG ACTCGTCCAA CATCAATACA ACCTATTAAT 3520
3521 TTCCCCTCGT CAAAAATAAG GTTATCAAGT GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG GCAAAAGCTT 3600
3601 ATGCATTTCT TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC AAACCGTTAT 3680
3681 TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA GGACAATTAC AAACAGGAAT CGAATGCAAC 3760
3761 CGGCGCAGGA ACACTGCCAG CGCATCAACA ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT 3840
3841 CCCGGGGATC GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA GGCATAAATT 3920
3921 CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACTCT 4000
4001 GGCGCATCGG GCTTCCCATA CAATCGATAG ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA 4080
4081 TAAATCAGCA TCCATGTTGG AATTTAATCG CGGCCTCGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA ACACCCCTTG 4160
4161 TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT 4240
4241 TGAGACACAA CGTGGCTTTC C                                                                   4261
```

6. The polynucleotide of Claim 4 from which the PAb sequence has been removed.

7. The plasmid vector of Claim 5 from which the PAb sequence has been removed.

8. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 ng and 100 mg of the polynucleotide of Claim 1 into the tissue of the vertebrate.

9. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into the tissue of a vertebrate the polynucleotide of Claim 1.

10. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 1 and a pharmaceutically acceptable carrier.

11. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 1 into the tissue of said primate and concurrently administering interleukin-12 parenterally.

12. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation an effector functions including lymphokine secretion specific to HCV antigens which comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 1.

13. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of Claim 1 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

14. A pharmaceutical composition comprising the polynucleotide of Claim 1.

15. A method of inducing an immune response comprising administering the polynucleotide of Claim 1 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant, recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

16. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 ng and 100 mg of the polynucleotide of Claim 2 into the tissue of the vertebrate.

17. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into the tissue of a vertebrate the polynucleotide of Claim 2.

18. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 2 and a pharmaceutically acceptable carrier.

19. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 2 into the tissue of said primate and concurrently administering interleukin 12 parenterally.

20. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation an effector functions including lymphokine secretion specific to HCV antigens which comprises exposing cells of a vertebrate <u>in vivo</u> to the polynucleotide of Claim 2.

21. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of Claim 2 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

22. A pharmaceutical composition comprising the polynucleotide of Claim 2.

23. A method of inducing an immune response comprising administering the polynucleotide of Claim 2 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant, recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

24. The vector of Claim 2 which is selected from V1Ra.HCV1CorePAb, Vtpa.HCV1CorePAb, VUb.HCV1CorePAb, V1Ra.HCV1Core, Vtpa.HCV1Core and VUb.HCV1Core.

25. A pharmaceutical composition comprising the vector of Claim 21.

26. The DNA sequence of Claim 1 selected from the group consisting of a nucleotide sequence shown in Figure 5, Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13.

```
          |   10       |   20       |   30       |   40       |   50       |   60       |   70       |   80
   1 GATATTGGCT ATTGGCCATT GCATACGTTG TATCCATATC ATAATATGTA CATTTATATT GGCTCATGTC CAACATTACC   80
  81 GCCATGTTGA CATTGATTAT TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT  160
 161 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC CATTGACGTC AATAATGACG  240
 241 TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC  320
 321 AGTACATCAA GTGTATCATA TGCCAAGTAC GCCCCCTATT GACGTCAATG ACGGTAAATG GCCCGCCTGG CATTATGCCC  400
 401 AGTACATGAC CTTATGGGAC TTTCCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA TTACCATGGT GATGCGGTTT  480
 481 TGGCAGTACA TCAATGGGCG TGGATAGCGG TTTGACTCAC GGGGATTTCC AAGTCTCCAC CCCATTGACG TCAATGGGAG  560
 561 TTTGTTTTGG CACCAAAATC AACGGGACTT TCCAAAATGT CGTAACAACT CCGCCCCATT GACGCAAATG GGCGGTAGGC  640
 641 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCGCCTGGA GACGCCATCC ACGCTGTTTT  720
 721 GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGCG GCCGGGAACG GTGCATTGGA ACGCGGATTC CCCGTGCCAA  800
 801 GAGTGACGTA AGTACCGCCT ATAGAGTCTA TAGGCCCACC CCCTTGGCTT CTTATGCATG CTATACTGTT TTTGGCTTGG  880
 881 GGTCTATACA CCCCCGCTTC CTCATGTTAT AGGTGATGGT ATAGCTTAGC CTATAGGTGT GGGTTATTGA CCATTATTGA  960
 961 CCACTCCCCT ATTGGTGACG ATACTTTCCA TTACTAATCC ATAACATGGC TCTTTGCCAC AACTCTCTTT ATTGGCTATA 1040
1041 TGCCAATACA CTGTCCTTCA GAGACTGACA CGGACTCTGT ATTTTTACAG GATGGGGTCT CATTTATTAT TTACAAATTC 1120
1121 ACATATACAA CACCACCGTC CCCAGTGCCC GCAGTTTTTA TTAAACATAA CGTGGGATCT CCACGCGAAT CTCGGGTACG 1200
1201 TGTTCCGGAC ATGGGCTCTT CTCCGGTAGC GGCGGAGCTT CTACATCCGA GCCCTGCTCC CATGCCTCCA GCGACTCATG 1280
1281 GTCGCTCGGC AGCTCCTTGC TCCTAACAGT GGAGGCCAGA CTTAGGCACA GCACGATGCC CACCACCACC AGTGTGCCGC 1360
1361 ACAAGGCCGT GGCGGTAGGG TATGTGTCTG AAAATGAGCT CGGGGAGCGG GCTTGCACCG CTGACGCATT TGGAAGACTT 1440
1441 AAGGCAGCGG CAGAAGAAGA TGCAGGCAGC TGAGTTGTTG TGTTCTGATA AGAGTCAGAG GTAACTCCCG TTGCGGTGCT 1520
1521 GTTAACGGTG GAGGGCAGTG TAGTCTGAGC AGTACTCGTT GCTGCCGCGC GCGCCACCAG ACATAATAGC TGACAGACTA 1600
1601 ACAGACTGTT CCTTTCCATG GGTCTTTTCT GCAGTCACCG TCCTTAGATC TAGGTACCAG ATATCAGAAT TCAGTCGACA 1680
1680 GCGGCCGCGA TCTGCTGTGC CTTCTAGTTG CCAGCCATCT GTTGTTTGCC CCTCCCCCGT GCCTTCCTTG ACCCTGGAAG 1760
1761 GTGCCACTCC CAGTGTCCTT TCCTAATAAA ATGAGGAAAT TGCATCGCAT TGTCTGAGTA GGTGTCATTC TATTCTGGGG 1840
1841 GCTGGGGTGG GGCAGCACAG CAAGGGGGAG GATTGGGAAG ACAATAGCAG GCATGCTGGG GATGCGGTGG GCTCTATGGG 1920
1921 TACGGCCGCA GCGGCCTTAA TTAAGGCCGC AGCGGCCGTA CCCAGGTGCT GAAGAATTGA CCCGGTTCCT CGACCCGTAA 2000
```

FIG.1A

```
2001 AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA AGTCAGAGGT 2080
2081 GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG 2160
2161 CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCAATGC TCACGCTGTA GGTATCTCAG 2240
2241 TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA 2320
2321 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG 2400
2401 AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 2480
2481 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG 2560
2561 GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGTGATCC 2640
2641 CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA ACTCATCGAG CATCAAATGA AACTGCAATT 2720
2721 TATTCATATC AGGATTATCA ATACCATATT TTTGAAAAAG CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC 2800
2801 CATAGGATGG CAAGATCCTG GTATCGGTCT GCGATTCCGA CTCGTCCAAC ATCAATACAA CCTATTAATT TCCCCTCGTC 2880
2881 AAAAATAAGG TTATCAAGTG AGAAATCACC ATGAGTGACG ACTGAATCCG GTGAGAATGG CAAAAGCTTA TGCATTTCTT 2860
2961 TCCAGACTTG TTCAACAGGC CAGCCATTAC GCTCGTCATC AAAATCACTC GCATCAACCA AACCGTTATT CATTCGTGAT 3040
3041 TGCGCCTGAG CGAGACGAAA TACGCGATCG CTGTTAAAAG GACAATTACA AACAGGAATC GAATGCAACC GGCGCAGGAA 3120
3121 CACTGCCAGC GCATCAACAA TATTTTCACC TGAATCAGGA TATTCTTCTA ATACCTGGAA TGCTGTTTTC CCGGGGATCG 3200
3201 CAGTGGTGAG TAACCATGCA TCATCAGGAG TACGGATAAA ATGCTTGATG GTCGGAAGAG GCATAAATTC CGTCAGCCAG 3280
3281 TTTAGTCTGA CCATCTCATC TGTAACATCA TTGGCAACGC TACCTTTGCC ATGTTTCAGA AACAACTCTG GCGCATCGGG 3360
3361 CTTCCCATAC AATCGATAGA TTGTCGCACC TGATTGCCCG ACATTATCGC GAGCCCATTT ATACCCATAT AAATCAGCAT 3440
3441 CCATGTTGGA ATTTAATCGC GGCCTCGAGC AAGACGTTTC CCGTTGAATA TGGCTCATAA CACCCCTTGT ATTACTGTTT 3520
3521 ATGTAAGCAG ACAGTTTTAT TGTTCATGAT GATATATTTT TATCTTGTGC AATGTAACAT CAGAGATTTT GAGACACAAC 3600
3601 GTGGCTTTCC                                                                             3610
             |  10      |  20      |  30      |  40      |  50      |  60      |  70      |  80
```

FIG.1B

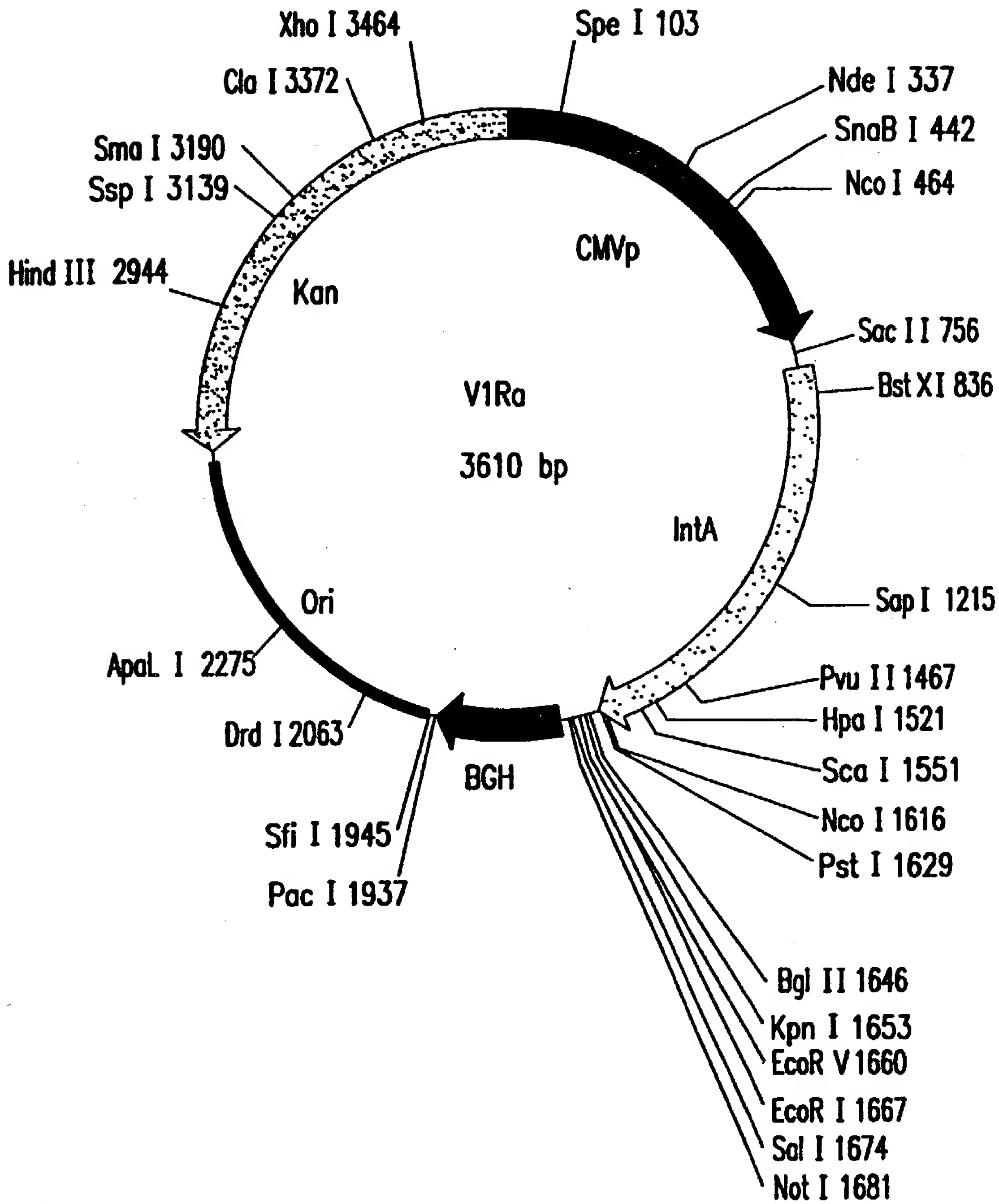
Xho I 3464
Spe I 103
Nde I 337
Cla I 3372
SnaB I 442
Sma I 3190
Ssp I 3139
Nco I 464
CMVp
Hind III 2944
Kan
Sac I I 756
Bst X I 836
V1Ra
3610 bp
IntA
Ori
Sap I 1215
ApaL I 2275
Pvu I I 1467
Drd I 2063
Hpa I 1521
Sca I 1551
BGH
Nco I 1616
Sfi I 1945
Pst I 1629
Pac I 1937
Bgl II 1646
Kpn I 1653
EcoR V 1660
EcoR I 1667
Sal I 1674
Not I 1681

FIG.2

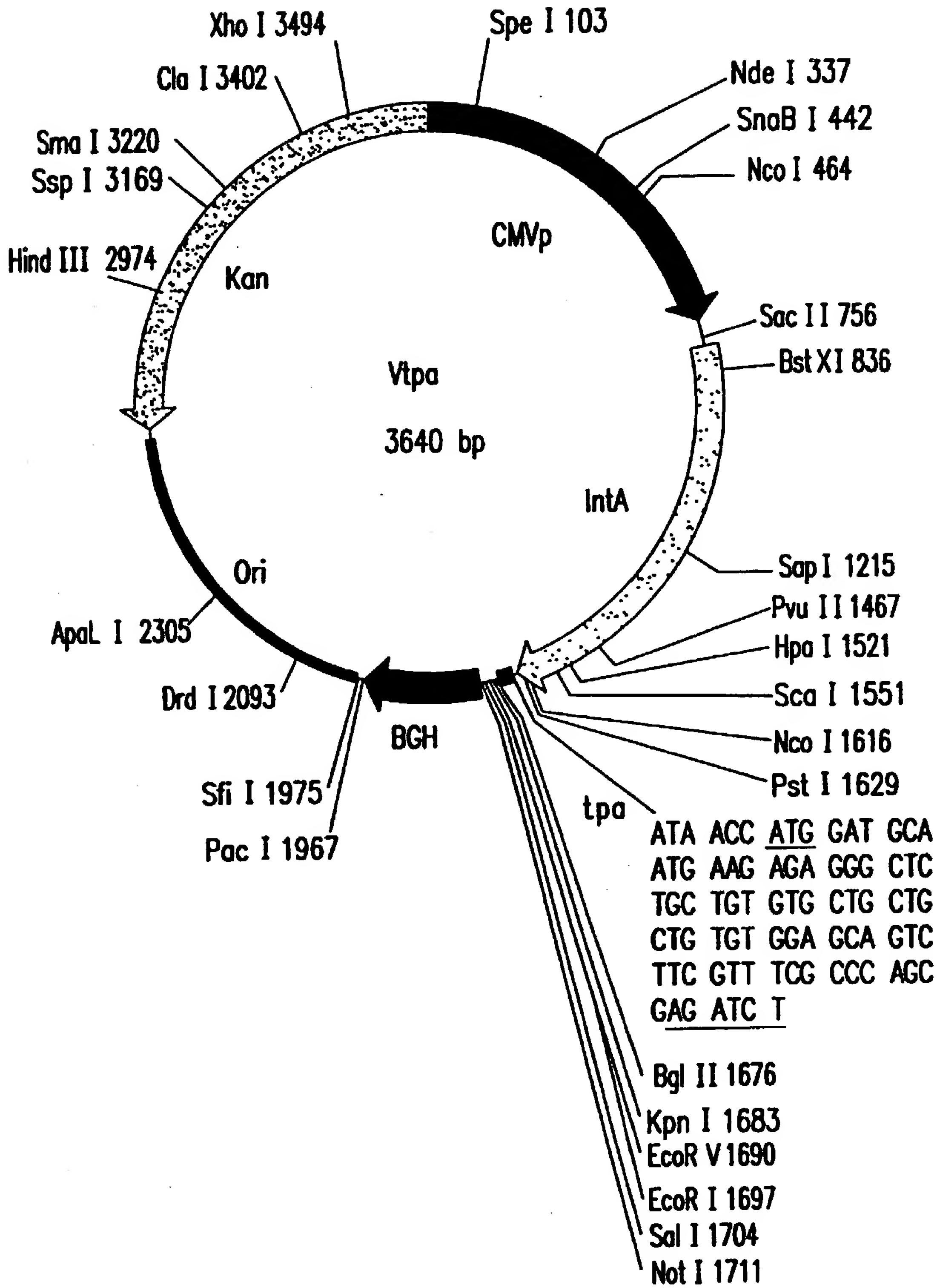
Xho I 3494
Spe I 103
Cla I 3402
Nde I 337
SnaB I 442
Sma I 3220
Nco I 464
Ssp I 3169
CMVp
Hind III 2974
Kan
Sac II 756
Bst XI 836
Vtpa
3640 bp
IntA
Sap I 1215
Ori
Pvu II 1467
ApaL I 2305
Hpa I 1521
Drd I 2093
Sca I 1551
Nco I 1616
BGH
Pst I 1629
Sfi I 1975
tpa
Pac I 1967
ATA ACC ATG GAT GCA
ATG AAG AGA GGG CTC
TGC TGT GTG CTG CTG
CTG TGT GGA GCA GTC
TTC GTT TCG CCC AGC
GAG ATC T
Bgl II 1676
Kpn I 1683
EcoR V 1690
EcoR I 1697
Sal I 1704
Not I 1711

FIG.3

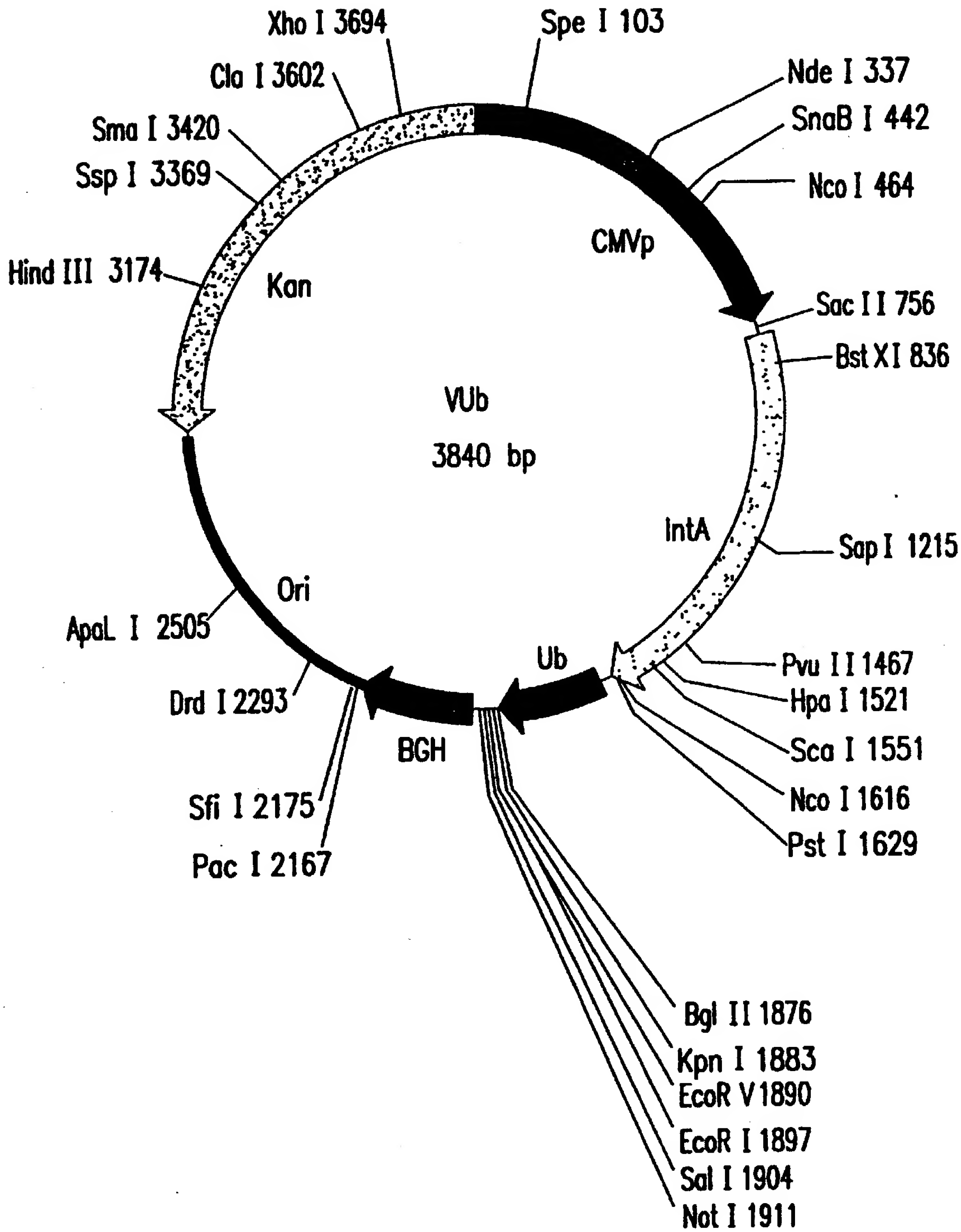
Xho I 3694
Spe I 103
Cla I 3602
Nde I 337
SnaB I 442
Sma I 3420
Ssp I 3369
Nco I 464
CMVp
Hind III 3174
Kan
Sac I I 756
Bst X I 836
VUb
3840 bp
IntA
Sap I 1215
Ori
ApaL I 2505
Ub
Pvu I I 1467
Hpa I 1521
Drd I 2293
BGH
Sca I 1551
Nco I 1616
Sfi I 2175
Pst I 1629
Pac I 2167
Bgl I I 1876
Kpn I 1883
EcoR V 1890
EcoR I 1897
Sal I 1904
Not I 1911

FIG.4

```
1/1                                     31/11
ATG AGC ACc AAc CCc AAg CCc CAg AGg AAg ACC AAg aGg AAC ACC AAC aGg aGg CCcCAG
Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr asn arg arg pro gln
61/21                                   91/31
GAt GTg AAG TTC CCt GGg GGa GGc CAG ATt GTg GGa GGg GTc TAC cTG cTG CCc aGg AGG
asp val lys phe pro gly gly gly gln ile val gly gly val tyr leu leu pro arg arg
121/41                                  151/51
GGC CCC AGG cTG GGg GTG aGg GCt ACc aGG AAG ACc TCt GAG aGG TCc CAg CCC aGg GGC
gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly
181/61                                  211/71
AGG aGg CAG CCc ATC CCC AAG GCc aGg aGG CCt GAG GGC cGc TCC TGG GCc CAG CCt GGc
arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly
241/81                                  271/81
TAC CCc TGG CCC CTg TAT GGC AAT GAa GGC TTt GGc TGG GCt GGc TGG CTG CTG TCC CCC
try pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro
301/101                                 331/111
aGg GGC TCc aGG CCc tcc TGG GGC CCC ACa GAC CCC aGG aGg aGG TCc aGg AAc cTG GGc
arg gly ser arg pro ser trp gly pro thr asp pro arg arg arg ser arg asn leu gly
361/121                                 391/131
AAg GTg ATt GAc ACC CTg ACc TGt GGC TTt GCt GAC CTg ATG GGc TAC ATC CCc CTg GTg
lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val
421/141                                 451/151
GGg GCt CCt GTg GGa GGg GTg GCt AGG GCt CTG GCt CAT GGg GTg AGG GTg CTG GAG GAt
gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp
481/161                                 511/171
GGG GTG AAC TAT GCt ACt GGc AAc cTG CCt GGc TGC TCc TTC TCc ATC TTC CTg CTG GCc
gly val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala
541/181                                 571/191
CTG CTc TCC TGC CTG ACa GTg CCt GCT TCT GCc
leu leu ser cys leu thr val pro ala ser ala
```

FIG. 5

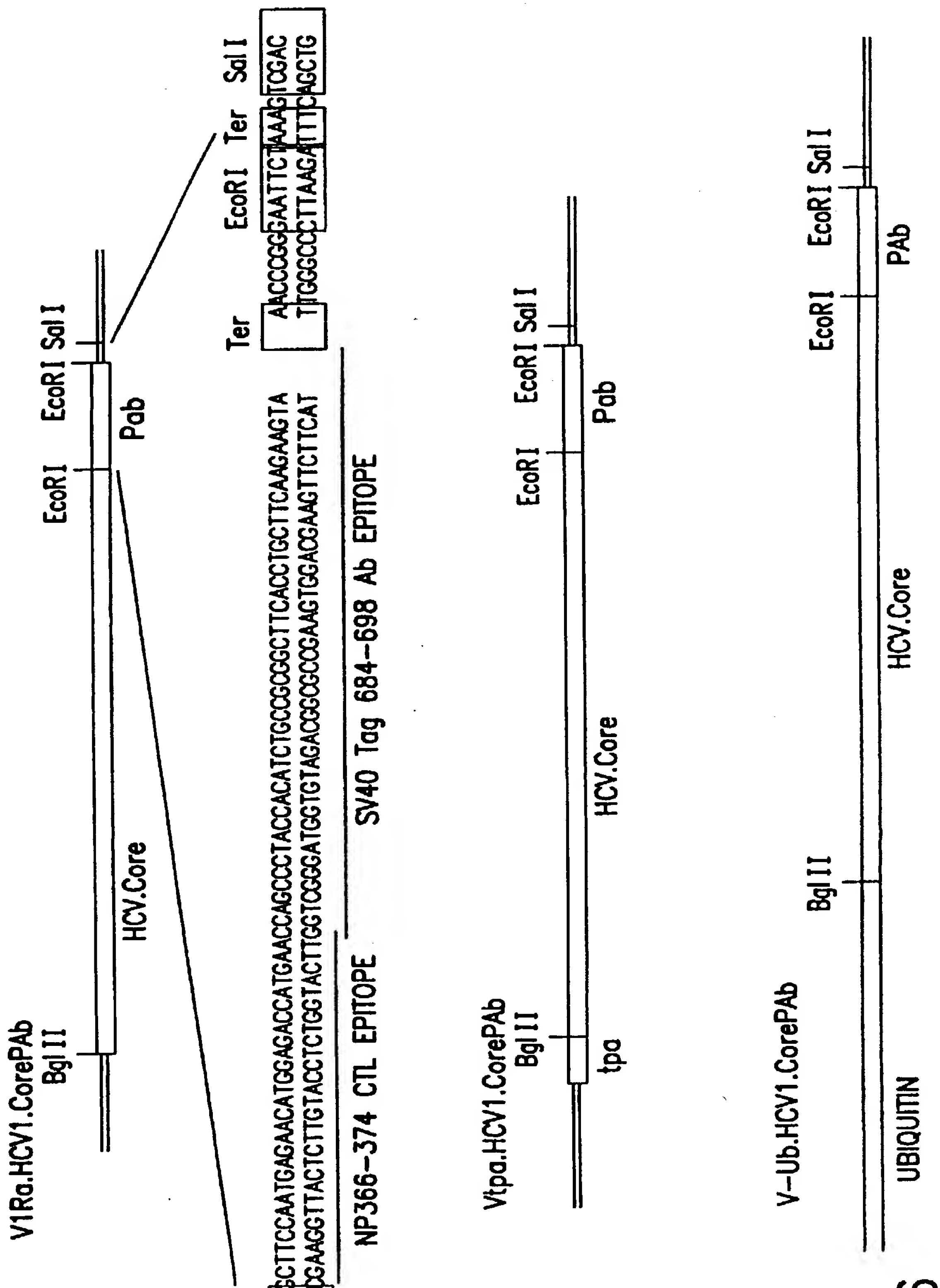
V1Ra.HCV1.CorePAb
Bgl II
HCV.Core
EcoRI
Pab
EcoRI
Sal I
EcoRI
GAATTCGCTTCCAATGAGAACATGGAGACCATGAACCAGCCCTACCACATCTGCGCGGCTTCACCTGCTTCAAGAAGTA
CTTAAGCGAAGGTTACTCTTGTACCTCTGGTACTTGGTCGGGATGGTGTAGACGCGCCGAAGTGGACGAAGTTCTTCAT
NP366-374 CTL EPITOPE
SV40 Tag 684-698 Ab EPITOPE
Ter
AACCCGGGAATTCTAAAGTCGAC
TTGGGCCCTTAAGATTTCAGCTG
EcoRI
Ter
Sal I
Vtpa.HCV1.CorePAb
Bgl II
tpa
HCV.Core
EcoRI
Pab
EcoRI
Sal I
V-Ub.HCV1.CorePAb
UBIQUITIN
Bgl II
HCV.Core
EcoRI
PAb
EcoRI
Sal I

FIG.6

```
1/1                                     31/11
ATG AGC ACG AAT CCT AAA CCT CAA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCa cAG
Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr asn arg arg pro gln
61/21                                   91/31
GAc GTc AAg TTC CCg GGC GGt GGt CAG ATC GTT GGt GGA GTT TAC TTC TTG CCG CGC AGG
asp val lys phe pro gly gly gly gln ile val gly gly val tyr leu leu pro arg arg
121/41                                  151/51
GGC CCC AGG TTG GGT GTG CGC GCG ACT aGG AAG ACT TCc GAG CGG TCG CAA CCT CGT GGa
gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly
181/61                                  211/71
AGG CGa CAG CCT ATC CCC AAG GCt CGc CGG cCC GAG GGc AGG TCC TGG GCT CAG CCC GGG
arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly
241/81                                  271/91
TAC CCt TGG CCc CTc TAt GGc AAT GAg GGC Ttc GGG TGG GCA GGa TGG CTC CTG TCC CCC
tyr pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro
301/101                                 331/111
CGC GGC TCT CGg CCt agT TGG GGC CCc AcT GAc CCC CGG CGt AGG TCG CGC AAT TTG GGT
arg gly ser arg pro ser trp gly pro thr asp pro arg arg arg ser arg asn leu gly
361/121                                 391/131
AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC GCC GAC CTC ATG GGg TAC ATC CCG CTC GTc
lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val
421/141                                 451/151
GGC GCC CCc GTA GGg GGC GTC GCC AGg GCC CTG GCG CAT GGC GTC AGG GtT cTG GAG GAC
gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp
481/161                                 511/171
GGG gtg AAC TAT GCA ACA GGG AAt tTg cCc GGT TGC TCT TTC TCT ATC TTC CTC cTG GCt
glu val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala
541/181                                 571/191
CTg CTg TCc TGC CTG ACC GTC CCA GCt TCT GCT
leu leu ser cys leu thr val pro ala ser ala
```

FIG. 7

TABLE 3
CODON UTILIZATION IN HUMAN PROTEIN-CODING SEQUENCES

a	b	c	d	e	f
F	UUU	68	0.35	193	4.5
	UUC	125	0.65		
L	UUA	20	0.05	445	10.4
	UUG	42	0.09		
	CUU	50	0.11		
	CUC	99	0.22		
	CUA	30	0.07		
	CUG	204	0.46		
I	AUU	28	0.23	123	2.9
	AUC	79	0.64		
	AUA	16	0.13		
M	AUG	77	1.00	77	1.8
V	GUU	35	0.13	266	6.2
	GUC	72	0.27		
	GUA	25	0.09		
	GUG	134	0.50		
S	UCU	59	0.17	349	8.1
	UCC	91	0.26		
	UCA	37	0.11		
	UCG	25	0.07		
	AGU	37	0.11		
	AGC	100	0.29		
P	CCU	51	0.24	212	4.9
	CCC	86	0.41		
	CCA	51	0.24		
	CCG	24	0.11		
T	ACU	47	0.20	238	5.6
	ACC	113	0.47		
	ACA	50	0.21		
	ACG	28	0.12		
A	GCU	91	0.31	298	7.0
	GCC	119	0.40		
	GCA	51	0.17		
	GCG	37	0.12		
Y	UAU	72	0.47	153	3.6
	UAC	81	0.53		
H	CAU	44	0.42	105	2.5
	CAC	61	0.58		
Q	CAA	50	0.26	192	4.5
	CAG	142	0.74		
N	AAU	51	0.34	148	3.5
	AAC	97	0.66		
K	AAA	137	0.45	303	7.0
	AAG	166	0.55		
D	GAU	79	0.38	209	4.9
	GAC	130	0.62		
E	GAA	125	0.40	311	7.3
	GAG	186	0.60		
C	UGU	44	0.30	147	3.4
	UGC	103	0.70		
W	UGG	56	1.00	56	1.3
R	CGV	19	0.09	215	5.0
	CGC	40	0.19		
	CGA	22	0.10		
	CGG	33	0.15		
	AGA	51	0.24		
	AGG	50	0.23		
G	GGU	36	0.15	245	5.7
	GGC	108	0.44		
	GGA	42	0.17		
	GGG	59	0.24		

TOTAL 4285 RESIDUES EXCLUDING
N-TERMINAL METHIONINE RESIDUES

FIG.8

```
1/1                                     31/11
atg TAT GAg GTG aGg AAt GTc TCt GGc GTc TAC CAT GTg ACc AAt GAC TGC TCC AAC TCc
M   Y   E   V   R   N   V   S   G   V   Y   H   V   T   N   D   C   S   N   S
61/21                                   91/31
tGc ATT GTc TAT GAG GCt GCt GAC ATG ATc ATG CAc ACC CCt GGc TGt GTg CCa TGt GTg
C   I   V   Y   E   A   A   D   M   I   M   H   T   P   G   C   V   P   C   V
121/41                                  151/71
aGG GAG GGc AAc TCC TCC aGg TGC TGG GTg GCc CTg ACc CCC ACc CTg GCt GCC AGG AAC
R   E   G   N   S   S   R   C   W   V   A   L   T   P   T   L   A   A   R   N
181/61                                  211/71
tcC tcC ATC CCC ACc ACc ACc ATc aGg aGg CAT GTg GAc cTG CTg GTg GGc GCt GCT GCc
S   S   I   P   T   T   T   I   R   R   H   V   D   L   L   V   G   A   A   A
241/81                                  271/91
CTg TGC TCt GCc ATG TAt GTG GGc GAc CTg TGT GGc TCT GTc TTC CTg GTg TCC CAg gTG
L   C   S   A   M   Y   V   G   D   L   C   G   S   V   F   L   V   S   Q   L
301/101                                 331/111
TTC ACC TTC TCc CCc aGg aGG TAT GAG ACt GTg CAG GAC TGC AAc TGC TCc CTg TAc CCt
F   T   F   S   P   R   R   Y   E   T   V   Q   D   C   N   C   S   L   Y   P
361/121                                 391/131
GGC CAt GTc TCt GGc CAC aGg ATG GCc TGG GAc ATG ATG ATG AAC TGG TCc CCc ACc ACt
G   H   V   S   G   H   R   M   A   W   D   M   M   M   N   W   S   P   T   T
421/141                                 451/151
GCC cTg GTG GTc TCc CAG cTg CTg aGG ATt CCc CAg GCt GTg GTG GAC ATG GTG TGT GGg
A   L   V   V   S   Q   L   L   R   I   P   Q   A   V   V   D   M   V   V   G
481/161                                 511/171
GCC CAC TGG GGc GTg CTG GCt GGC CTg GCC TAC TAc TCC ATG GTG GGc AAC TGG GCc AAG
A   H   W   G   V   L   A   G   L   A   Y   Y   S   M   V   G   N   W   A   K
541/181                                 571/191
GTg cTG ATT GTG ATG CTg CTg TTT GCt GGC GTg GAt GGc taa
V   L   I   V   M   L   L   F   A   G   V   D   G   *
```

FIG. 9

```
1/1                                         31/11
atg ACc ACC TAt GTc TCt GTG GGc CAT GCc tcC CAG ACC ACC aGG aGg GTg GCc TCC TTC
M   T   T   Y   V   S   V   G   H   A   S   Q   T   T   R   R   V   A   S   F
61/21                                       91/31
TTC tcc CCT GGC TCt GCc CAG AAg ATC CAg CTg GTg AAC ACC AAt GGC tcc TGG CAC ATC
F   S   P   G   S   A   Q   K   I   Q   L   V   N   T   N   G   S   W   H   I
121/41                                      151/51
AAC AGG ACT GCC CTG AAt TGC AAt GAG TCC ATC AAC ATC GGc TTC TTT GCt GCc CTG TTC
N   R   T   A   L   N   C   N   E   S   I   N   T   G   F   F   A   A   L   F
181/61                                      211/71
TAt GTg AAG AAG TTC AAC TCc TCT GGc TGC TCt GAG aGg ATG GCc tct TGc aGg CCC ATT
Y   V   K   K   F   N   S   S   G   C   S   E   R   M   A   S   C   R   P   I
241/81                                      271/91
GAC AGG TTt GCc CAg GGc TGG GGc CCC ATC ACC CAT GCT GAG TCc aGg tcC TCt GAC CAg
D   R   F   A   Q   G   W   G   P   I   T   H   A   E   S   R   S   S   D   Q
301/101                                     331/111
AGG CCa TAC TGC TGG CAC TAt GCc CCc CAg CCa TGT GGc ATt GTG CCt GCc cTG CAt GTc
R   P   Y   C   W   H   Y   A   P   Q   P   C   G   I   V   P   A   L   H   V
361/121                                     391/131
TGt GGc CCt GTc TAc TGc TTC ACC CCa tcC CCT GTg GTg GTG GGc ACg ACt GAC aGg TTt
C   G   P   V   Y   C   F   T   P   S   P   V   V   V   G   T   T   D   R   F
421/141                                     451/151
GGC GTg CCC ACc TAc AAC TGG GGc GAC AAT GAG ACt GAt GTG CTg CTg CTG AAC AAC ACc
G   V   P   T   Y   N   W   G   D   N   E   T   D   V   L   L   L   N   N   T
481/161                                     511/171
aGG CCc CCc CAg GGc AAC TGG TTt GGC TGc ACc TGG ATG AAC tcC ACt GGc TTC ACC AAG
R   P   P   Q   G   N   W   F   G   C   T   W   M   N   S   T   G   F   T   K
```

FIG.10A

```
541/181                                 571/191
ACc TGt GGc GGC CCC CCa TGc AAc ATt GGc GGc GCt GGC AAC AAC ACC cTG ACC TGC CCC
T   C   G   G   P   P   C   N   I   G   G   A   G   N   N   T   L   T   C   P
601/201                                 631/211
ACt GAc TGC TTC aGG AAG CAt CCt GAG GCC ACc TAC ACC AAG TGt GGC TCt GGc CCa TGG
T   D   C   F   R   K   H   P   E   A   T   Y   T   K   C   G   S   G   P   W
661/221                                 691/231
cTG ACc CCc AGG TGC ATG GTg GAC TAC CCa TAC AGg CTg TGG CAC TAC CCa TGC ACc TTC
L   T   P   R   C   M   V   D   Y   P   Y   R   L   W   H   Y   P   C   T   F
721/241                                 751/251
AAC TTc ACC ATC TTc AAG ATC AGG ATG TAT GTG GGc GGc GTG GAG CAC AGG CTg AAt GCt
N   F   T   I   F   K   I   R   M   Y   V   G   G   V   E   H   R   L   N   A
781/261                                 811/271
GCc TGC AAC TGG ACc aGg GGc GAG aGg TGc AAC ATt GAG GAC AGG GAc AGG TCt GAG CTg
A   C   N   W   T   R   G   E   R   C   N   I   E   D   R   D   R   S   E   L
841/281                                 871/291
tcC CCc CTG CTg CTG TCc ACc ACt GAG TGG CAG ATc CTg CCa TGc TCC TTc ACC ACC CTg
S   P   L   L   L   S   T   T   E   W   Q   I   L   P   C   S   F   T   T   L
901/301                                 931/311
CCt GCc CTG TCC ACT GGc cTG ATC CAt CTg CAt CAG AAC ATt GTG GAt GTG CAg TAC CTG
P   A   L   S   T   G   L   I   H   L   H   Q   N   I   V   D   V   Q   Y   L
961/321                                 991/331
TAt GGc GTg GGc TCt GCt GTg GTc TCC ATT GTG ATC AAg TGG GAG TAt GTg CTG CTG CTg
Y   G   V   G   S   A   V   V   S   I   V   I   K   W   E   Y   V   L   L   L
1021/341
TTC CTg CTg CTG GCt GAt GCc taa
F   L   L   L   A   D   A   *
```

FIG.10B

```
1/1                                     31/11
atg TAT GAg GTG aGg AAt GTc TCt GGc GTc TAC CAT GTg ACc AAt GAC TGC TCC AAC TCc
M   Y   E   V   R   N   V   S   G   V   Y   H   V   T   N   D   C   S   N   S
61/21                                   91/31
tGc ATT GTc TAT GAG GCt GCt GAC ATG ATc ATG CAc ACC CCt GGc TGt GTg CCa TGt GTg
C   I   V   Y   E   A   A   D   M   I   M   H   T   P   G   C   V   P   C   V
121/41                                  151/51
aGG GAG GGc AAc TCC TCC aGg TGC TGG GTg GCc CTg ACc CCC ACc CTg GCt GCC AGG AAC
R   E   G   N   S   S   R   C   W   V   A   L   T   P   T   L   A   A   R   N
181/61                                  211/71
tcC tcC ATC CCC ACc ACc ACc ATc aGg aGg CAT GTg GAc cTG CTg GTg GGc GCt GCT GCc
S   S   I   P   T   T   T   I   R   R   H   V   D   L   L   V   G   A   A   A
241/81                                  271/91
CTg TGC TCt GCc ATG TAt GTG GGc GAc CTg TGT GGc TCT GTc TTC CTg GTg TCC CAg cTG
L   C   S   A   M   Y   V   G   D   L   C   G   S   V   F   L   V   S   Q   L
301/101                                 331/111
TTC ACC TTC TCc CCc aGg aGG TAT GAG ACt GTg CAG GAC TGC AAc TGC TCc CTg TAc CCt
F   T   F   S   P   R   R   Y   E   T   V   Q   D   C   N   C   S   L   Y   P
361/121                                 391/131
GGC CAt GTc TCt GGc CAC aGg ATG GCc TGG GAc ATG ATG ATG AAC TGG TCc CCc ACc ACt
G   H   V   S   G   H   R   M   A   W   D   M   M   M   N   W   S   P   T   T
421/141                                 451/151
GCC cTg GTG GTc TCc CAG cTg CTg aGG ATt CCc CAg GCt GTg GTG GAC ATG GTG GTG GGc
A   L   V   V   S   Q   L   L   R   I   P   Q   A   V   V   D   M   V   V   G
481/161                                 511/171
GCC CAC TGG GGc GTg CTG GCt GGC CTg GCC TAC TAc TCC ATG GTG GGc AAC TGG GCc AAG
A   H   W   G   V   L   A   G   L   A   Y   Y   S   M   V   G   N   W   A   K
```

FIG.11A

```
541/181                                 571/191
GTg cTG ATT GTG ATG CTg CTg TTT GCt GGC GTg GAt GGc ACc ACC TAt GTc TCt GTG GGc
V   L   I   V   M   L   L   F   A   G   V   D   G   T   T   Y   V   S   V   G
601/201                                 631/211
CAT GCc tcC CAG ACC ACC aGG aGg GTg GCc TCC TTC TTC tcc CCT GGC TCt GCc CAG AAg
H   A   S   Q   T   T   R   R   V   A   S   F   F   S   P   G   S   A   Q   K
661/221                                 691/231
ATC CAg CTg GTg AAC ACC AAt GGC tcc TGG CAC ATC AAC AGG ACT GCC CTG AAt TGC AAt
I   Q   L   V   N   T   N   G   S   W   H   I   N   R   T   A   L   N   C   N
721/241                                 751/251
GAG TCC ATC AAC ACT GGc TTC TTT GCt GCc CTG TTC TAt GTg AAG AAG TTC AAC TCc TCT
E   S   I   N   T   G   F   F   A   A   L   F   Y   V   K   K   F   N   S   S
781/261                                 811/271
GGc TGC TCt GAG aGg ATG GCc tct TGc aGg CCC ATT GAC AGG TTt GCc CAg GGc TGG GGc
G   C   S   E   R   M   A   S   C   R   P   I   D   R   F   A   Q   G   W   G
841/281                                 871/291
CCC ATC ACC CAT GCT GAG TCc aGg tcC TCt GAC CAg AGG CCa TAC TGC TGG CAC TAt GCc
P   I   T   H   A   E   S   R   S   S   D   Q   R   P   Y   C   W   H   Y   A
901/301                                 931/311
CCc CAg CCa TGT GGc ATt GTG CCt GCc cTG CAt GTc TGt GGc CCt GTc TAc TGc TTC ACC
P   Q   P   C   G   I   V   P   A   L   H   V   C   G   P   V   Y   C   F   T
961/321                                 991/331
CCa tcC CCT GTg GTg GTG GGc ACc ACt GAC aGg TTt GGC GTg CCC ACc TAc AAC TGG GGc
P   S   P   V   V   V   G   T   T   D   R   F   G   V   P   T   Y   N   W   G
1021/341                                1051/351
GAC AAT GAG ACt GAt GTG CTg CTg CTg AAC AAC ACc aGG CCc CCc CAg GGc AAC TGG TTe
D   N   E   T   D   V   L   L   L   N   N   T   R   P   P   Q   G   N   W   F
```

FIG.11B

```
1081/361                                1111/37
GGC TGc ACc TGG ATG AAC tcC ACt GGc TTC ACC AAG ACc TGt GGc GGC CCC CCa TGc AAc
G   C   T   W   M   N   S   T   G   F   T   K   T   C   G   G   P   P   C   N
1141/381                                1171/391
ATt GGc GGc GCt GGC AAC AAC ACC cTG ACC TGC CCC ACt GAc TGC TTC aGG AAG CAt CCt
I   G   G   A   G   N   N   T   L   T   C   P   T   D   C   F   R   K   H   P
1201/401                                1231/411
GAG GCC ACc TAC ACC AAG TGt GGC TCt GGc CCa TGG cTG ACc CCc AGG TGC ATG GTg GAC
E   A   T   Y   T   K   C   G   S   G   P   W   L   T   P   R   C   M   V   D
1261/421                                1291/431
TAC CCa TAC AGg CTg TGG CAC TAC CCa TGC ACc TTC AAC TTc ACC ATC TTc AAG ATC AGG
Y   P   Y   R   L   W   H   Y   P   C   T   F   N   F   T   I   F   K   I   R
1321/441                                1351/451
ATG TAT GTG GGc GGc GTG GAG CAC AGG CTg AAt GCt GCc TGC AAC TGG ACc aGg GGc GAG
M   Y   V   G   G   V   E   H   R   L   N   A   A   C   N   W   T   R   G   E
1381/461                                1411/471
aGg TGc AAC ATg GAG CAC AGG GAc AGG TCt GAG CTg tcC CCc CTG CTg CTG TCc ACc ACt
R   C   N   I   E   D   R   D   R   S   E   L   S   P   L   L   L   S   T   T
1441/481                                1471/491
GAG TGG CAG ATc CTg CCa TGc TCC TTc ACC ACC CTg CCt GCc CTG TCC ACT GGc cTG ATC
E   W   Q   I   L   P   C   S   F   T   T   L   P   A   L   S   T   G   L   I
1501/501                                1531/511
CAt CTg CAt CAG AAC ATt GTG GAt GTG CAg TAC CTG TAt GGc GTg GGc TCt GCt GTg GTc
H   L   H   Q   N   I   V   D   V   Q   Y   L   Y   G   V   G   S   A   V   V
1561/521                                1591/531
TCC ATT GTG ATC AAg TGG GAG TAt GTg CTG CTG CTg TTC CTg CTg CTG GCt GAt GCc taa
S   I   V   I   K   W   E   Y   V   L   L   L   F   L   L   L   A   D   A   *
```

FIG.11C

```
1/1                                     31/11
atg TCt GGc TCc TGG CTg AGG GAT GTc TGG GAC TGG ATc TGC ACt GTG cTG ACT GAC TCC
M   S   G   S   W   L   R   D   V   W   D   W   I   C   T   V   L   T   D   F
61/21                                   91/31
AAG ACC TGG CTg CAt TCC AAG CTg CTG CCc aGG CTG CCt GGc GAC CCa TTC TTC TCc TGc
K   T   W   L   H   S   K   L   L   P   R   L   P   G   D   P   F   F   S   C
121/41                                  151/51
CAg aGg GGc TAC AGG GGc GTc TGG aGG GGc GAT GGC GTg ATG CAg ACC ACC TGC CCa TGT
Q   R   G   Y   R   G   V   W   R   G   D   G   V   M   Q   T   T   C   P   C
181/61                                  211/71
GGc GCc CAG ATC ACt GGc CAT GTg AAg AAt GGc TCC ATG AGG ATt GTg GGc CCc AAg ACC
G   A   Q   I   T   G   H   V   K   N   G   S   M   R   I   V   G   P   K   T
241/81                                  271/91
TGc tcC AAC ACc TGG CAt GGc ACc TTC CCC ATC AAt GCc TAC ACC ACt GGc CCa TGC ACc
C   S   N   T   W   H   G   T   F   P   I   N   A   Y   T   T   G   P   C   T
301/101                                 331/111
CCa TCC CCt GCc CCc AAC TAc TCC AGG GCc CTG TGG aG GTG GCT GCT GAG GAG TAT GTG
P   S   P   A   P   N   Y   S   R   A   L   W   R   V   A   A   E   Y   V   V
361/121                                 391/131
GAg GTg ACc aGG GTG GGc GAc TTC CAC TAt GTG ACt GGC ATG ACC ACT GAC AAt GTg AAg
E   V   T   R   V   G   D   F   H   Y   V   T   G   M   T   T   D   N   V   K
421/141                                 451/151
TGC CCa TGC CAG GTg CCt GCC CCt GAg TTC TTC ACt GAg GTG GAT GGc GTG aGG cTG CAC
C   P   C   Q   V   P   A   P   E   F   F   T   E   V   D   G   V   R   L   H
481/161                                 511/171
AGG TAt GCc CCt GCc TGC AAg CCc CTg CTg aGG GAT GAG GTg ACc TTC CAG GTg GGc CTg
R   Y   A   P   A   C   K   P   L   L   R   D   E   V   T   F   Q   V   G   L
```

FIG.12A

```
541/181                                 571/191
AAC CAg TTC CCt GTg GGc TCc CAG CTg CCa TGT GAG CCt GAg CCt GAT GTg ACt GTG CTg
N   Q   F   P   V   G   S   Q   L   P   C   E   P   E   P   D   V   T   V   L
601/201                                 631/211
ACc TCC ATG CTg ACt GAg CCa TCC CAC ATc ACt GCt GAG ACt GCc AAG aGg AGG cTG GCC
T   S   M   L   T   E   P   S   H   I   T   A   E   T   A   K   R   R   L   A
661/221                                 691/231
AGg GGc TCc CCt CCa TCC cTG GCC tcC TCc TGCc tcC CAG cTG TCT GCt CCa TCc cTG
R   G   S   P   P   S   L   A   S   S   S   A   S   Q   L   S   A   P   S   L
721/241                                 751/251
AAG GCc ACc TGC ACc ACC aGg CAT GAC TCC CCt GAt GCt GAC CTg ATt GAG GCC AAC CTg
K   A   T   C   T   T   R   H   D   S   P   D   A   D   L   I   E   A   N   L
781/261                                 811/271
CTG TGG aGG CAG GAG ATG GGC GGc AAC ATC ACC aGG GTG GAG TCt GAG AAc AAG GTg GTg
L   W   R   Q   E   M   G   G   N   I   T   R   V   E   S   E   N   K   V   V
841/281                                 871/291
ATc CTg GAC TCc TTT GAg CCc CTg aGG GCt GAG GAG GAT GAG AGG GAg GTc TCt GTG GCt
I   L   D   S   F   E   P   L   R   A   E   E   D   E   R   E   V   S   V   A
901/301                                 931/311
GCt GAG ATC CTg aGG AAg tcC AGG AAG TTC CCC CCt GCc cTG CCC ATc TGG GCg aGg CCa
A   E   I   L   R   K   S   R   K   F   P   P   A   L   P   I   W   A   R   P
961/321                                 991/331
tcC TAC AAC CCa CCc CTg CTg GAG TCc TGG AAG GAC CCt GAC TAt GTg CCc CCt GTG GTg
S   Y   N   P   P   L   L   E   S   W   K   D   P   D   Y   V   P   P   V   V
1021/381                                1051/371
CAt GGc TGC CCc CTG CCc CCc ACC ATG GCC CCa CCc ATc CCc CCa CCc aGG AGG AAG AGG
H   G   C   P   L   P   P   T   M   A   P   P   I   P   P   P   R   R   K   R
```

FIG.12B

```
1081/361                                1111/371
ACt GTg GTg CTG ACt GAg TCC ACt GTc TCc TCT GCC cTG GCT GAG CTg GCC ACc AAG ACC
T   V   V   L   T   E   S   T   V   S   S   A   L   A   E   L   A   T   K   T
1141/381                                1171/391
TTC GGC tcC TCt GGc TCc TCt GCt GTg GAC tct GGC ACt GCc ACG GCC CCc CCT GAC CAG
F   G   S   S   G   S   S   A   V   D   S   G   T   A   T   A   P   P   D   Q
1201/401                                1231/411
CCa TCt GAT GAt GGc GAC AGg GGc TCt GAt GAT GAG TCc TAC TCC TCC ATG CCC CCC CTg
P   S   D   D   G   D   R   G   S   D   D   E   S   Y   S   S   M   P   P   L
1261/421                                1291/431
GAG GGc GAG CCt GGc GAC CCt GAc CTg tct GAt GGc TCc TGG TCc ACt GTc tct GAG GAG
E   G   E   P   G   D   P   D   L   S   D   G   S   W   S   T   V   S   E   E
1321/441
GCc tct GAG GAt GTg GCc TGC TGC TCc taa
A   S   E   D   V   A   C   C   S   *
```

FIG.12C

```
1/1                                     31/11
ATG TCc TAC ACc TGG ACt GGC GCC CTg ATC ACc CCa TGt GCt GCt GAG GAG tcC AAG CTG
M   S   Y   T   W   T   G   A   L   I   T   P   C   A   A   E   E   S   K   L
61/21                                   91/31
CCC ATC AAc CCc cTG tcC AAC TCc cTG CTG aGg CAt CAC AAC ATG GTc TAt GCC ACc ACc
P   I   N   P   L   S   N   S   L   L   R   H   H   N   M   V   Y   A   T   T
121/41                                  151/51
TCc aGg tct GCt GGC CTg aGG CAG AAg AAG GTg ACC TTT GAC AGg CTG CAt GTg CCt GAt
S   R   S   A   G   L   R   Q   K   K   V   T   F   D   R   L   H   V   P   C
181/61                                  211/71
GAC CAc TAC aGG GAt GTG CTg AAG GAG ATG AAG GCc AAG GCc TCC ACt GTg AAG GCg AAg
D   H   Y   R   D   V   L   K   E   M   K   A   K   A   S   T   V   K   A   K
241/81                                  271/91
CTg CTg TCT GTg GAg GAg GCC TGC AAG CTG ACc CCt CCc CAC TCt GCC AGg TCC AAg TTT
L   L   S   V   E   E   A   C   K   L   T   P   P   H   S   A   R   S   K   F
301/101                                 331/111
GGC TAT GGc GCc AAG GAt GTg aGG AAC CTg TCC tcC AAG GCt GTg AAC CAC ATC CAC TCt
G   Y   G   A   K   D   V   R   N   L   S   S   K   A   V   N   H   I   H   S
361/121                                 391/131
GTc TGG AAG GAC cTG CTG GAg GAC ACT GAg ACc CCc ATT GAC ACc ACC ATC ATG GCc AAg
V   W   K   D   L   L   E   D   T   E   T   P   I   D   T   T   I   M   A   K
421/141                                 451/151
AAT GAG GTc TTC TGT GTg CAg CCt GAG AAg GGc GGC aGg AAG CCt GCC aGg CTg ATt GTc
N   E   V   F   C   V   Q   P   E   K   G   G   R   K   P   A   R   L   I   V
481/161                                 511/171
TTC CCt GAg CTg GGc GTg aGg GTG TGt GAG AAg ATG GCC CTg TAt GAt GTG GTc TCC ACc
F   P   E   L   G   V   R   V   C   E   K   M   A   L   Y   D   V   V   S   T
```

FIG.13A

```
541/181                                 571/191
CTg CCc CAG GCt GTG ATG GGC TCC TCc TAt GGc TTC CAG TAC TCc CCT GGc CAG aGG GTg
L   P   Q   A   V   M   G   S   S   Y   G   F   Q   Y   S   P   G   Q   R   V
601/201                                 631/211
GAG TTC CTG GTG AAT GCC TGG AAg TCc AAG AAg AAC CCc ATG GGC TTt GCc TAc TGC ACC
E   F   L   V   N   A   W   K   S   K   K   N   P   M   G   F   A   Y   C   T
661/221                                 691/231
aGg TGc TTT GAC TCc ACt GTg ACT GAG tcT GAc ATC aGg GTg GAG GAG TCc ATc TAC CAg
R   C   F   D   S   T   V   T   E   S   D   I   R   V   E   E   S   I   Y   Q
721/241                                 751/251
TGc TGT GAC cTG GCt CCt GAG GCC AGg CAG GTg ATc AGG TCc CTg ACt GAG aGG CTg TAc
C   C   D   L   A   P   E   A   R   Q   V   I   R   S   L   T   E   R   L   Y
781/261                                 811/271
ATt GGc GGC CCC CTG ACc AAc TCc AAg GGc CAG AAC TGt GGc TAc aGg aGG TGC aGg GCC
I   G   G   P   L   T   N   S   K   G   Q   N   C   G   Y   R   R   C   R   A
841/281                                 871/291
tct GGc GTG CTG ACc ACT AAC TGt GGc AAc ACC CTg ACc TGc TAC cTG AAG GCC TCT GCt
S   G   V   L   T   T   N   C   G   N   T   L   T   C   Y   L   K   A   S   A
901/301                                 931/311
GCt TGc aGg GCT GCc AAG CTg CAt GAC TGC ACc ATG CTg GTc TGt GGc GAt GAC CTg GTg
A   C   R   A   A   K   L   H   D   C   T   M   L   V   C   G   D   D   L   V
961/321                                 991/331
GTg ATC TGT GAg tct GCt GGc ACC CAG GAG GAt GCt GCc tcC CTg aGg GTc TTC ACt GAG
V   I   C   E   S   A   G   T   Q   E   D   A   A   S   L   R   V   F   T   E
1021/341                                1051/351
GCc ATG ACc AGG TAC TCT GCC CCc CCt GGc GAC CCt CCC CAg CCt GAg TAt GAC cTG GAG
A   M   T   R   Y   S   A   P   P   G   D   P   P   Q   P   E   Y   D   L   E
```

FIG.13B

```
1081/361                                1111/371
cTg ATc ACc TCc TGC TCC TCC AAt GTc TCt GTg GCc CAt GAt GCc TCT GGC AAG aGG GTc
L   I   T   S   C   S   S   N   V   S   V   A   H   D   A   S   G   K   R   V
1141/381                                1171/391
TAC TAC CTg ACc aGg GAC CCC ACC ACC CCc CTg GCc AGG GCt GCc TGG GAg ACt GCc AGg
Y   Y   L   T   R   D   P   T   T   P   L   A   R   A   A   W   E   T   A   R
1201/401                                1231/411
CAC ACc CCt GTg AAC TCC TGG CTg GGC AAC ATC ATC ATG TAt GCc CCC ACC CTG TGG GCc
H   T   P   V   N   S   W   L   G   N   I   I   M   Y   A   P   T   L   W   A
1261/421                                1291/431
AGG ATG ATc CTG ATG ACC CAC TTC TTC TCC ATC CTg CTg GCc CAG GAG CAg CTg GAg AAg
R   M   I   L   M   T   H   F   F   S   I   L   L   A   Q   E   Q   L   E   K
1321/441                                1351/451
GCC CTG GGc TGc CAG ATt TAt GGc GCC ACc TAC TTC ATT GAg CCc CTg GAC CTg CCc CAG
A   L   G   C   Q   I   Y   G   A   T   Y   F   I   E   P   L   D   L   P   Q
1381/461                                1411/471
ATC ATc CAG aGg CTg CAt GGc CTg tct GCc TTc TCc CTg CAC tcc TAC TCc CCt GGc GAg
I   I   Q   R   L   H   G   L   S   A   F   S   L   H   S   Y   S   P   G   E
1441/481                                1471/491
ATC AAc AGG GTG GCc TCc TGC CTg AGG AAg CTg GGc GTg CCc CCC cTG aGg GTg TGG AGg
I   N   R   V   A   S   C   L   R   K   L   G   V   P   P   L   R   V   W   R
1501/501                                1531/511
GAc aGG GCC AGg tct GTg aGg GCc AAG CTg CTG TCC CAG GGc GGc AGG GCt GCC ACc TGT
H   R   A   R   S   V   R   A   K   L   L   S   Q   G   G   R   A   A   T   C
1561/521                                1591/531
GGC AAG TAC CTg TTC AAC TGG GCt GTG AGG ACC AAG CTg AAg CTg ACc CCc ATc CCt GCT
G   K   Y   L   F   N   W   A   V   R   T   K   L   K   L   T   P   I   P   A
```

FIG.13C

```
1621/541                                1651/551
GCc TCC CAG cTG GAC cTg TCt GGC TGG TTt GTg GCT GGc TAC tct GGc GGc GAC ATc TAc
A   S   Q   L   D   L   S   G   W   F   V   A   G   Y   S   G   G   D   I   Y
1681/561                                1711/571
CAC tcC CTG TCc aGg GCC aGg CCC aGg TGG TTC ATG TGG TGC CTg CTg CTg CTg TCT GTg
H   S   L   S   R   A   R   P   R   W   F   M   W   C   L   L   L   L   S   V
1741                                    1771/591
GGc GTg GGC ATC TAC CTG CTg CCC AAC aGG TGA
G   V   G   I   Y   L   L   P   N   R   *
```

FIG.13D

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/09884

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61N 43/04; C12Q 1/68; C12N 15/00; C07H 21/02; A61K 39/00
US CL : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN, MEDLINE, BIOSIS, EMBASE, CAPLUS, WPIDS, SCISEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Selby et al. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. Journal of General Virology. 1993. Vol. 74, pages 1103-1113, see entire document.	1-3
X	Bukh et al. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc. Natl. Acad. Sci. August 1994. Vol. 91, pages 8239-8243, see entire document.	1-3
Y	Lathe. Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data Theoretical and Practical Considerations. J. Mol. Biol. 1985. Vol. 183, pages 1-12, see entire document.	1-3

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 AUGUST 1997	11 SEP 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
ANDREW WANG
Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/09884

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Grantham et al. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Research. 1981. Vol. 9, No. 1, pages r43-r74, see entire document.	1-3
A, P	Ide et al. Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. Gene. December 1996. Vol. 182, pages 203-211, see entire document.	1-3, 8-26
X	US 5,514,539 A (BUKH et al.) 07 May 1996, see entire document.	1-3, 8-26

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/09884

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 4-7
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

The enclosed copy of claims 4, 5 were not legible and claims 6, 7 depend on those claims.

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.